EXTRACTING EMOTIONS IN SONGS USING ACOUSTIC AND LYRICAL FEATURES

Rahul Duggal[1], Shampa Chakraverty[2], Jyoti Narang[3]

[1]IIIT Delhi, [2]NSIT Delhi, [3]Desein Dot Ventures

[1]rahulduggal2608@gmail.com, [2]apmahs.nsit@gmail.com, [3]jyotinarang.nsit@gmail.com

ABSTRACT: *In this paper, we formulate the task of predicting song emotions as a multi-label classification problem using a combination of musical and lyrical features. We use the Geneva Emotion Music Scale (GEMS-9), an emotion classification system specifically designed to capture emotions in songs. We invited online users from varied walks of life to annotate songs compiled from top-10 charts of a popular music website, with felt emotions. For every song, we extract a set of 13 musical features including high level features such as Acousticness, Danceability and Instrumentalness. We extract its lyrical features with topic modeling by employing Latent Dirichlet Allocation. For classification, we train a per emotion category one-versus-all SVM classifier and demonstrate how lyrics and acoustic features when used together, improve the classification accuracy. Results show that the Precision of the combined features approach exceeds that of acoustic-only classification by 8.9% and that of lyrics-only classification by 9.4%.*

1. INTRODUCTION

Music is a means of expression that mankind has evolved over centuries to connect people by evoking in them, a common feeling. Indeed, this intrinsic connect between music and mood/behavior is the primary driver for playing ambient background music in hotel lobbies and restaurants, in trains and buses during travel, in museums and in hospitals [1]. A recent study in [2] shows how music can affect activity in brain structures that are known to be crucially responsible for emotions. These indicate a natural connect between music and emotions.

In today's technological landscape, the task of predicting emotions in songs finds several applications such as in streaming music, music information retrieval and music recommender systems. According to[1], 164 billion songs were streamed in the US alone in 2014. The sheer number of songs releasing everyday makes it the need of the hour to classify, retrieve and recommend relevant songs in a reliable manner. These factors have motivated several research initiatives and our work is an effort in this direction. The main contributions of this paper are summarized below:

1. We use a combination of several high level musical and lyrical features to assess the consolidated impact of these two mutually complimentary components of a song.
2. We use GEMS9 emotion classification [15], which has been specifically designed for capturing emotions in music and is therefore suited for the task of emotion prediction in songs. Past approaches have used generic emotion classification schemes.
3. We apply topic modeling using LDA [3] to capture the lyrical features of songs. This affords the advantage of gauging the quality of the lyrical features by observing the constituent words of each topic.

[1] 2014 nielsen music u.s. report, http://www.nielsen.com/content/dam/corporate/us/en/public%20factsheets/Soundscan/nielsen-2014-year-end-music-report-us.pdf

4. We compiled a database of labeled songs by inviting people from diverse backgrounds to annotate songs through our public domain website[2] which is open to the research community.

2. PRIOR WORK

Prior works have used acoustics features [4, 6, 7, 11], lyrical features [5, 8, 9, 10], and both kinds of features [12]. In [12], the authors have shown that when lyrics are combined with acoustic features, the classification accuracy of emotion prediction increases since the words of a song also contribute towards evoking emotions. This motivates us to adopt both these aspects for our work. However, in contrast with [12] that uses low level acoustic features such as rhythm and timbre, we experiment with a richer variety of 13 musical features combined with lyrical features.

Prior works have used generic emotion taxonomies such as Thayer's model [13], Russell's and Tellegen-Watson Clark's model of emotion [14]. While these classification approaches are useful in a general emotion detection setting, they fall short of capturing the spectrum of emotions felt by a music listener. A recent breakthrough in this direction came through the Geneva Emotion Music Scales (GEMS9) classification which was specifically developed to capture a majority of emotions that are evoked through music [15]. Thus in our study, we use the GEMS9 model.

Various classification methods such as regression [6], fuzzy based classification [10, 11], Multiclass classification [8], and multi-label classification [4] have been applied for song emotion detection. The GEMS9 classification describes the emotion space using 9 broad emotions. Since the emotions are not mutually exclusive, multiple emotions may be tagged to a single song. Thus multilabel classification is the natural choice to model the prediction problem. In comparison with [4] which uses a total of 6 individual labels, our dataset contains songs, each of which is tagged with any combination of 9 labels.

3. EMOTION CLASSIFICATION MODEL, DATASET AND FEATURES

GEMS9 Emotion Classification: Though GEMS originally proposed 45 emotion labels, we use a condensed form of the scale known as GEMS9. Table 1 describes the spectrum of emotions that are included in GEMS9.

Dataset Preparation: Every year the popular music themed website *www.billboard.com* releases a genre-wise top 100 chart. For our dataset, we used the top 10 songs in four genres – Rock, Pop, R&B and country from the period between 2011 and 2015. We were able to sample 183 unique songs since there were some songs common to genres. We made a website[2] for the task of labeling the songs with emotions. After listening, the user can chose to label the song with a maximum of 3 emotions.

Emotional category	Explanation	Emotional category	Explanation
Calmness	Relaxation, serenity, meditativeness	Solemnity	Feeling of transcendence, inspiration. Thrills
Tenderness	Sensuality, affect, feeling of love	Power	Feeling strong, heroic, triumphant, energetic
Sadness	Depressed, sorrowful	Joyful activation	Feels like dancing, bouncy feeling, animated, amused
Tension	Nervous, impatient, irritated	Nostalgia	Dreamy, melancholic, sentimental feelings
Amazement	Feeling of wonder and happiness		

Table 1: GEMS-9 Emotion Classification

[2]www.*emoticator.in*

Song Name	Artist	Emotion Profile
Shut Up And Dance	Walk The Moon	Joyful activation
Somebody That I Used To Know	Gotye Featuring Kimbra	Nostalgia, Sadness
The Monster	Eminem	Power
Thinking Out Loud	Ed Sheeran	Tenderness, Calmness

Table 2: Emotion profiles of some songs in our dataset

The website[2] has been gathering votes since July 2016 from a diverse set of users consisting of different age groups, socio cultural backgrounds and professions. The average and median number of upvotes per song in our dataset are 8.55 and 8 respectively. Moreover every song was annotated by at least 5 people. This mass participation based compilation is in contrast with the existing public dataset in [4] where only three music experts labeled the entire dataset. The final emotion profile for a song was determined by considering only those emotions which gathered votes exceeding a certain threshold, determined experimentally. The emotion profiles of some songs in our dataset are shown in Table 2.

Feature Set: We employ a combination of lyrical and acoustic features described below.

Lyrical features: To determine the lyrical features of a song, we employ the well-known topic modeling technique, Latent Dirichlet Allocation (LDA). The key idea is that any document (song), is synthesized from a small collection of topics and each word in it is probabilistically attributable to these latent topics. Our motivation to use LDA is that every song may be considered to be composed of a mixture of latent topics in a way that discriminates them according to the emotions they convey. It is interesting to note that we did observe meaningful topics emerging from our dataset as shown in Table 3. Topic 12 has a predominance of words that occur in many joyful songs. Topic 35 clearly contains explicit words in many rap songs.

Topic #	Top Words
12	0.107*diamond + 0.073*bright + 0.060*sky + 0.042*shine + 0.041*beautiful + 0.037*ready + 0.036*fall +0.034*shine +0.030*eye + 0.026*have
35	0.139*bitch + 0.087*fuck + 0.076*nigga + 0.064*love + 0.037*niggas + 0.036*ball + 0.036*bout + 0.033*money + 0.029*long + 0.028*day

Table 3: Interesting topics returned by LDA.

Feature	Description	Feature	Description
Acousticness	Confidence measure of acoustic content	Liveness	Whether song was performed live
Danceability	Confidence measure of how danceable the song is.	Loudness	Loudness of the track in dB
Duration_ms	Length of the song in milliseconds	Mode	Modality (major/minor) of the song
Energy	Perceptual measure of intensity and activity.	Speechiness	Measure of spoken words in the song
Instrumentalness	Measure of 'No vocals' in the track.	Tempo	An estimate of Beats per minute
Key	The key signature of the song	Time_signature	number of beats per bar
		Valence	Measure of musical positivity conveyed by the track.

Table 4: Acoustic features

Acoustic Features: We used a combination of 13 basic and high level acoustic features given in Table 4. We extracted these features by employing a web based API provided by spotify[3] . The high level features are composed by considering several low level features in a composite manner. For example, *Acousticness* is composed of tempo, *rhythm, stability, beat strength* and *overall regularity*. Likewise, *Energy* is built from dynamic *range, perceived loudness, timbre, onset rate* and *general entropy*.

4. PROCESS FLOW

Figure 2 gives an overview of the flow of various processes that our proposed emotion detection system undergoes. The procedure starts off on the lyrical front by correcting spellings as lyricists often intentionally misspell words to mimic the intended pronunciation, such as 'layin' for 'laying'. The removal of stop words is a standard procedure in many language processing tasks since words like 'a', 'the' etc. do not convey any emotion. Lemmatization is an important step because the meaning of text is contained predominantly in concept terms, *i.e.* nouns, verbs and adjectives. Also, the inflected forms of a word like play, playing, played convey the same conceptual meaning.

For lyrical features, we use the latent topic weights found by applying LDA on each song in the database. There are two parameters which affect the classification accuracy in this step. These are (i) the number of topics and (ii) the parameter *alpha* [3]. We found empirically that 40 topics give the highest classification accuracy. *Alpha* controls the sparsity of the latent topics; lower *alpha* results in sparser representations. *Alpha* was empirically set to 1.5.

Normalization is necessary since the range of features varies in magnitude, and features with larger ranges would otherwise dominate. We use PCA as a feature reduction step. With our dataset, this resulted in a reduction in number of features from original 53 to 45. The classifier used was a one-versus-rest rbf kernel SVM classifier [16]. We use 10-fold cross validation to train the classifier and evaluate its performance.

5. EXPERIMENTS AND RESULTS

In order to assess the classification quality, we used micro-averaged precision, recall and F1 metrics. The micro-averaged scores are amortized over all classes of emotions. It predominantly reflects the classifier's performance on those classes whose instances occur most frequently. These represent the most commonly occurring emotion classes in the dataset.

$$\text{Micro-Precision (P)} = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + FP_i} \quad (1)$$

$$\text{Micro-Recall (R)} = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + FN_i} \quad (2)$$

$$\text{Micro-F1 score} = \frac{2 \times P \times R}{P + R} \quad (3)$$

Where n is the number of emotion classes (=9 for GEM9 based classification), TP_i , FP_i and FN_i refer to the number of True positives, False positives and False negatives respectively, for the ith emotion class.

[3] https://developer.spotify.com/web-api/get-several-audio-features/

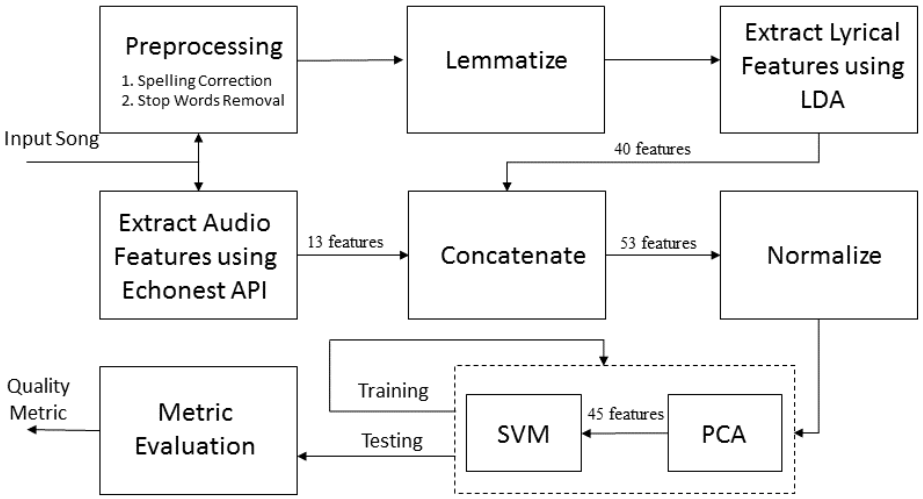

Figure 2: Overview of the proposed method for song emotion prediction

TYPE	METRIC	ACOUSTIC ONLY (13 FEATURES)	LYRICAL ONLY (40 FEATURES)	ACOUSTIC + LYRICAL (53 FEATURES)	ACOUSTIC + LYRICAL + PCA (45 FEATURES)
MICRO	Precision	0.594	0.585	**0.647**	0.64
	Recall	0.336	0.325	0.351	**0.356**
	f1 score	0.43	0.416	0.455	**0.456**

Table 5: Results on various metrics

We experimented on the compilation of annotated songs according to the process flow in Figure 2. We found that were variations in cross validated results in different iterations. Therefore, we first consolidated the results over 100 iterations of 10 fold cross validation and then reported their average. The results based on various classification metrics are tabulated in Table 5.

It can be seen from Table 5 that the combined lyrical plus acoustic feature model clearly outperforms the single modality models. The corresponding values obtained after applying PCA are comparable. The combined model with audio + lyrics + PCA features yields a micro-averaged Precision of 0.64 which is an improvement of 8.9% over audio only model and 9.4% over the lyrics only model. It yields a Recall measure of 0.356 which is 4.46% higher than that obtained with audio only features and 9.53% higher than lyrics only features. The final F1 score of 0.456 for the composite features case is 5.8% greater than audio only model and 9.6% greater than the lyrics only model.

6. CONCLUSION AND FUTURE WORK

We have proposed a song emotion prediction system based on the GEM-9 emotion classification system which is designed specifically for the purpose of describing emotions in songs. Further, we have included a comprehensive set of high level acoustic features and lyrical features. We experimented on a carefully compiled database labeled with emotions voted by a large number of online listeners through our public domain website. Results demonstrate that lyrical features obtained through LDA combined with high level acoustic features result in better quality of classification in terms of precision, recall and F1 measure. In order to further improve the accuracy, we are examining other features relevant for emotion detection in songs such as the sentiments and semantics of words in lyrics.

REFERENCES

[1] Ryu, K., & Jang, S. S. (2007). The effect of environmental perceptions on behavioral intentions through emotions: The case of upscale restaurants.*Journal of Hospitality & Tourism Research*, *31*(1), 56-72.

[2] Koelsch, S. (2014). Brain correlates of music-evoked emotions. Nature Reviews Neuroscience, 15(3), 170-180. Koelsch, S. (2014). Brain correlates of music-evoked emotions. Nature Reviews Neuroscience, 15(3), 170-180.

[3] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation.*Journal of machine Learning research*, *3*(Jan), 993-1022.

[4] Trohidis, K., Tsoumakas, G., Kalliris, G., &Vlahavas, I. P. (2008, September). Multi-Label Classification of Music into Emotions. In *ISMIR*(Vol. 8, pp. 325-330).

[5] Yang, D., & Lee, W. S. (2009, December). Music emotion identification from lyrics. In *Multimedia, 2009. ISM'09. 11th IEEE International Symposium on*(pp. 624-629). IEEE.

[6] Yang, Y. H., Lin, Y. C., Su, Y. F., & Chen, H. H. (2007, July). Music emotion classification: A regression approach. In *2007 IEEE International Conference on Multimedia and Expo* (pp. 208-211). IEEE.

[7] Song, Y., Dixon, S., & Pearce, M. (2012, October). Evaluation of Musical Features for Emotion Classification. In *ISMIR* (pp. 523-528).

[8] Kim, M., &Kwon, H. C. (2011, November). Lyrics-based emotion classification using feature selection by partial syntactic analysis. In *2011 IEEE 23rd International Conference on Tools with Artificial Intelligence* (pp. 960-964). IEEE.

[9] He, H., Jin, J., Xiong, Y., Chen, B., Sun, W., & Zhao, L. (2008, December). Language feature mining for music emotion classification via supervised learning from lyrics. In *International Symposium on Intelligence Computation and Applications* (pp. 426-435). Springer Berlin Heidelberg.

[10] Hu, Y., Chen, X., & Yang, D. (2009, August). Lyric-based Song Emotion Detection with Affective Lexicon and Fuzzy Clustering Method. In *ISMIR*(pp. 123-128).

[11] Yang, Y. H., Liu, C. C., & Chen, H. H. (2006, October). Music emotion classification: a fuzzy approach. In *Proceedings of the 14th ACM international conference on Multimedia* (pp. 81-84). ACM.

[12] Yang, Y. H., Lin, Y. C., Cheng, H. T., Liao, I. B., Ho, Y. C., & Chen, H. H. (2008, December). Toward multi-modal music emotion classification. In*Pacific-Rim Conference on Multimedia* (pp. 70-79). Springer Berlin Heidelberg.

[13]R.E. Thayer. The Biopsychology of Mood and Arousal. Oxford University Press, 1989.

[14] A. Tellegen, D. Watson, and L.A. Clark. On the dimen-sional and hierarchical structure of affect. Psychological Science, 10(4):297–303, July 1999.

[15] Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement.*Emotion*, *8*(4), 494.

[16] Rifkin, R., &Klautau, A. (2004). In defense of one-vs-all classification.*Journal of machine learning research*, *5*(Jan), 101-141.